DTIC FILE COPY

INTEGRATING SYNTAX, SEMANTICS, AND DISCOURSE
DARPA NATURAL LANGUAGE UNDERSTANDING PROGRAM

R&D STATUS REPORT

AD-A194 098   Unisys/Defense Systems

ARPA ORDER NUMBER: 5262
PROGRAM CODE NO. NR 049-602 dated 10 August 1984 (433)
CONTRACTOR: Unisys Defense Systems   Paoli, PA.   CONTRACT AMOUNT: $872,833
CONTRACT NO: N00014-85-C-0012
EFFECTIVE DATE OF CONTRACT: 4/29/85   EXPIRATION DATE OF CONTRACT: 4/28/89
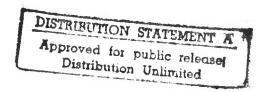PRINCIPAL INVESTIGATOR: Dr. Lynette Hirschman   PHONE NO. (215) 648-7554
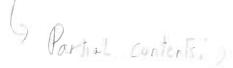
SHORT TITLE OF WORK: DARPA Natural Language Understanding Program

REPORTING PERIOD: - 8/1/87-10/31/87

DTIC
ELECTE
MAR 3 0 1988
S
D

88   3   30   001

Partial contents;

1. Description of Progress

A report was submitted to the ONR describing our research activities under the DARPA contract for the 1987 fiscal year.

1.1. Syntax/Pragmatics Interaction

A mechanism was added to allow syntactically bound null arguments of verbs to be coindexed with an elided element. For example, in the sentence *refused to heave to*, the subject of the infinitive *heave to* is coindidexed with the elided subject of the main verb *refuse*. When the reference resolution component determines the appropriate referent of the refusal event, this referent is automatically resolved to be the subject of the infinitive as well. Previously this coindexing was only available for full noun phrases or pronouns.

1.2. Semantics

The semantic interpreter has been extended to handle left past participial modifiers in noun phrases, such as *sheared connecting pin* as well as right participial modifiers of noun phrases, such as *lube oil saturated with metallic particles*. The semantic interpreter has also been reorganized to eliminate use of the internal database for storing semantic rules. This will make development easier and should also speed lookup.

The treatment of the phenomenon of "semantic raising", which was developed for the RAINFORM domain is being extended to the CASREPS domain. This provides a domain-independent semantic treatment to find the implicit arguments of nominalizations occurring as objects in constructions like *sac received high usage*, where the subject of *receive*, the sac, is also an argument of the nominalization *usage*. Several semantic raising verb rules have been added to the CASEPS rules. A new aspectual operator has been added to the time component to handle the temporal properties of these verb + object combinations.

1.3. Work on Multiple Domains

PUNDIT's code was reorganized so that domain-dependent and domain-independent code now reside in distinct files and distinct directories. Separate directories and files for each domain makes it possible to work simultaneously on multiple domains. The utilities which create executable images from this code (system compile files) were updated to be compatible with the new directory structure. These utilities have been modified to make it easier for individual developers to tailor the working system for their own needs. They have also been considerably modified in order to automate both the generation of domain-specific images and the testing of their output. For example, we can now automatically run large sets of messages with a single command.

1.4. Environment

In order to facilitate the use of the new Xwindows interface, system images are now maintained under Quintus Prolog 2.0 as well as 1.5. Using the newer Prolog also required minor changes to the system compile utilities.

1.5. New Domains

We have continued our collaboration on and pursuit of custom applications for natural language processing in several distinct domains. In addition to the Trident messages described below, we have also investigated medical and intelligence message domains.

1.5.1. Naval Domain

In the naval domain, we have investigated message-processing applied to Trident messages; these messages are collected under a contract originally awarded to Sperry. We are evaluating the suitability of adding this

domain to the DARPA contract work. As part of our support, we provided a 2-day training seminar to give technical staff on this contract a detailed introduction to PUNDIT and to the issues involved in porting PUNDIT to a new message domain. Our new staff members also attended this seminar. We prepared a booklet of seminar materials which will continue to be used for training and general documentation purposes.

Also as part of the Trident effort we are developing a version of the system to process a selected set of messages. We have already added the new vocabulary to the lexicon, and have extended the shapes component which parses part numbers so that it will recognize and use features of these strings, such as those distinguishing part numbers from procedure identifiers. We have also begun designing the semantic rules for this domain, and will be adding a capacity to recognize the discourse structure of these messages for use in generating and updating a message database.

2. Change in Key Personnel

Shirley Steele, who holds a Ph.D. in Communication Disorders from the Unversity of Texas at Dallas, joined us on September 14 from AT&T Bell Labs at Murray Hill, NJ. She will be initiating our speech effort and has begun exploring the possibility of collaborating with the MIT Speech Labs.

Carl Weir, who holds a Ph.D. in Linguistics from the University of Texas at Austin, started work on September 8. He formerly worked at MCC, and his expertise is in semantics and pragmatics.

3. Summary of Substantive Information from Meetings and Conferences

3.1. Professional Meetings Attended

John Dowding and Lynette Hirschman attended the 1987 IEEE Logic Programming Symposium in San Francisco (August 31-September 4).

Marcia Linebarger attended a conference on "Planning for Future Research: Directions for the Next Decade" sponsored by the Northeast Artificial Intelligence Consortium (NAIC) at the Minnowbrook Conference Center, Blue Mountain Lake, NY. September 20 through 23.

Shirley Steele and Lynette Hirschman attended a meeting of Darpa speech contractors held at BBN, October 13-15.

Catherine Ball presented a paper called "Natural Language Processing: An Overview" at the USE, Inc. conference in Montreal on October 15.

3.2. Papers and Presentations

Catherine Ball presented a paper called "Natural Language Processing: An Overview" at the USE, Inc. conference in Montreal on October 15.

Deborah Dahl presented an invited colloquium to the Department of Computer Science at Penn State entitled "Integration of Semantics and Pragmatics in the Computational Analysis of Nominalizations".

John Dowding and Lynette Hirschman presented a paper entitled "A Dynamic Translator for Rule Pruning in Restriction Grammar" at the Second International Workshop on Natural Language Understanding and Logic Programming, in Vancouver, B.C., in August.

Lynette Hirschman gave a tutorial on Natural Language and Logic Programming at the 1987 IEEE Logic Programming Symposium.

A paper by Francois Lang and Lynette Hirschman entitled "Improved Portability and Parsing Through Interactive Acquisition of Semantic Information" has been accepted for presentation at the Applied ACL conference, which will take place in February in Austin, Texas.

0. **Problems Expected or Anticipated**

   None.

1. **Action Required by the Government**

   None.

2. **Fiscal Status**

(1) Amount currently provided on contract:

   $ 872,833 (funded) $1,704,901 (contract value)

(2) Expenditures and commitments to date:

   $ 743,173

(3) Funds required to complete work:

   $ 129,660

Accesion For		
NTIS CRA&I	☑	
DTIC TAB	☐	
Unannounced	☐	
Justification		
By per ltr		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		